(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 117292696 B
(45)授权公告日 2024.03.12

(21)申请号 202311301500.4
(22)申请日 2023.10.08
(65)同一申请的已公布的文献号
申请公布号 CN 117292696 A
(43)申请公布日 2023.12.26
(73)专利权人 合肥工业大学
地址 230009 安徽省合肥市包河区屯溪路193号
(72)发明人 乔亚涛 苏兆品 岳峰 张国富
(74)专利代理机构 北京久诚知识产权代理事务所(特殊普通合伙) 11542
专利代理师 肖柏红
(51)Int.Cl.
G10L 19/018(2013.01)
G10L 25/30(2013.01)
G10L 19/16(2013.01)

(56)对比文件
CN 104867496 A,2015.08.26
CN 109587372 A,2019.04.05
CN 111640444 A,2020.09.08
CN 113077377 A,2021.07.06
CN 113965659 A,2022.01.21
EP 4064095 A1,2022.09.28
US 2019189111 A1,2019.06.20
US 2021192019 A1,2021.06.24
沈朝勇.基于差分进化的音频智能隐写算法研究.《中国优秀硕士学位论文全文数据库 信息科技辑(月刊) 电信技术》.2023,(第05期),全文.
苏兆品,张羚,张国富,岳峰.基于多特征融合和BiLSTM的语音隐写检测算法.《电子学报》.2023,第51卷(第5期),全文.

审查员 严雪莹

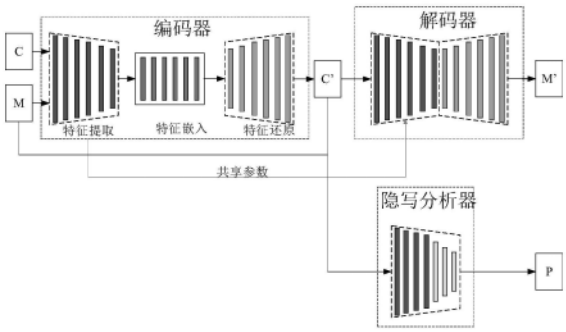
权利要求书2页 说明书10页 附图5页

(54)发明名称

端到端音频隐写方法、系统、存储介质及电子设备

(57)摘要

本发明提供一种端到端音频隐写方法、系统、存储介质及电子设备,涉及音频处理技术领域。本发明通过循环自编码器进行生成对抗网络预训练,确定编码器中特征提取模块和特征还原模块的参数,且基于生成对抗网络框架设计了端到端的隐写算法,不仅避免了因为STFT不匹配导致的秘密信息提取失败问题,同时取消了载体音频的修改向量,使编码器直接生成载密音频,从而达到降低模型的训练难度并提高模型性能的目的,有效解决了现有的音频隐写方法稳定性差的技术问题。



CN 117292696 B

1.一种端到端音频隐写方法,其特征在于,采用生成对抗网络预先构建编码器和隐写分析器,根据编码器预先构建解码器,所述端到端音频隐写方法包括:

S1、获取秘密音频和载体音频,并通过预先训练的编码器对秘密音频和载体音频进行处理,输出载密音频;

S2、通过解码器对载密音频进行解密处理,输出秘密音频的估计音频;

其中,通过循环自编码器进行生成对抗网络预训练,确定编码器中特征提取模块和特征还原模块的参数;

所述特征提取模块用于提取并联合秘密音频时间依赖特征和载体音频时间依赖特征,得到时间依赖特征;

所述编码器还包括特征嵌入模块,所述特征嵌入模块用于高维展开时间依赖特征,并进行秘密特征的嵌入,得到嵌入秘密音频特征的载密融合特征;

所述特征还原模块用于对载密融合特征进还原,输出载密音频;

所述解码器中包括第二特征提取模块和第二特征还原模块,所述第二特征提取模块通过共享编码器中特征提取模块的网络参数,第二特征提取模块的结构、参数和编码器中特征提取模块中的结构、参数保持一致。

2.如权利要求1所述的端到端音频隐写方法,其特征在于,所述特征提取模块包括依次连通的6个Convblock和1个拼接层,其中,第一个Convblock的输入通道数为1,输出通道数为64和卷积核大小为3×3,第二个Convblock的输入通道数为64,输出通道数为64和卷积核大小为1×3,第三个Convblock的输入通道数为64,输出通道数为128和卷积核大小为1×3,第四个Convblock的输入通道数为128,输出通道数为128和卷积核大小为1×3,第五个Convblock的输入通道数为128,输出通道数为128和卷积核大小为1×3;第六个Convblock的输入通道数为256,输出通道数为256和卷积核大小为1×3。

3.如权利要求2所述的端到端音频隐写方法,其特征在于,所述特征嵌入模块包括依次连通的8个mixblock,8个mixblock的卷积核大小均为3×3,其中,第一个mixblock的输入通道数为512,输出通道数为576,第二个mixblock的输入通道数为576,输出通道数为640,第三个mixblock的输入通道数为640,输出通道数为768,第四个mixblock的输入通道数为768,输出通道数为1024,第五个mixblock的输入通道数为1024,输出通道数为768,第六个mixblock的输入通道数为768,输出通道数为576,第七个mixblock的输入通道数为576,输出通道数为512,第四个mixblock的输入通道数为512,输出通道数为256。

4.如权利要求1所述的端到端音频隐写方法,其特征在于,所述特征还原模块包括依次连通的6个Transblock,其中,前五个Transblock的卷积核为1×3,第六个Transblock的卷积核为3×3,第一个Transblock的输入通道数为256,输出通道数为256,第二个Transblock的输入通道数为256,输出通道数为128,第三个Transblock的输入通道数为128,输出通道数为128,第四个Transblock的输入通道数为128,输出通道数为64,第五个Transblock的输入通道数为64,输出通道数为64,第六个Transblock的输入通道数为64,输出通道数为1。

5.如权利要求1所述的端到端音频隐写方法,其特征在于,所述隐写分析器包括依次连通的4个Convblock、3个Linearblock和一层softmax层。

6.如权利要求1～5任一所述的端到端音频隐写方法,其特征在于,所述编码器、隐写分析器和解码器训练过程中的损失函数包括:

$L_S=x\log(S(C))+(1-x)\log(1-S(C'))$

$L_D=\text{Distortion}(M,M')$

$L_E=\lambda_1(\text{Distortion}(C,C'))+\lambda_2 L_S+\lambda_3 L_D$

$$\text{Distortion}(C,C')=1-\frac{\sum y_i' y_i}{\sum(y_i'^2+y_i^2)-\sum y_i' y_i}$$

其中,L_E表示编码器的损失;L_D表示解码器的损失;L_S表示隐写分析器的损失;λ_1、λ_2、λ_3分别表示编码器、隐写分析器、解码器的损失所占的权重系数;S(C)表示被隐写分析器S识别为载体音频的概率,S(C')表示被识别为载密音频的概率;x表示隐写分析器的标签,将编码器产生的载密音频标签为1,将原始的载体音频标签为0;$y=\{y_1,y_2,\cdots,y_i,\cdots,y_n\}$表示时域载体音频,$y'=\{y'_1,y'_2,\cdots y'_i\cdots,y'_n\}$表示时域载密音频。

7.一种端到端音频隐写系统,其特征在于,采用生成对抗网络预先构建编码器和隐写分析器,根据编码器预先构建解码器,该端到端音频隐写系统包括:

加密模块,用于获取秘密音频和载体音频,并通过预先训练的编码器对秘密音频和载体音频进行处理,输出载密音频;

解码模块,用于通过解码器对载密音频进行解密处理,输出秘密音频的估计音频;

其中,通过循环自编码器进行生成对抗网络预训练,确定编码器中特征提取模块和特征还原模块的参数;

所述特征提取模块用于提取并联合秘密音频时间依赖特征和载体音频时间依赖特征,得到时间依赖特征;

所述编码器还包括特征嵌入模块,所述特征嵌入模块用于高维展开时间依赖特征,并进行秘密特征的嵌入,得到嵌入秘密音频特征的载密融合特征;

所述特征还原模块用于对载密融合特征进还原,输出载密音频;

所述解码器中包括第二特征提取模块和第二特征还原模块,所述第二特征提取模块通过共享编码器中特征提取模块的网络参数,第二特征提取模块的结构、参数和编码器中特征提取模块中的结构、参数保持一致。

8.一种计算机可读存储介质,其特征在于,其存储用于端到端音频隐写的计算机程序,其中,所述计算机程序使得计算机执行如权利要求1~6任一所述的端到端音频隐写方法。

9.一种电子设备,其特征在于,包括:

一个或多个处理器,存储器,以及一个或多个程序,其中所述一个或多个程序被存储在所述存储器中,并且被配置成由所述一个或多个处理器执行,所述程序包括用于执行如权利要求1~6任一所述的端到端音频隐写方法。

端到端音频隐写方法、系统、存储介质及电子设备

技术领域

[0001] 本发明涉及音频处理技术领域,具体涉及一种端到端音频隐写方法、系统、存储介质及电子设备。

背景技术

[0002] 随着因特网的普及、信息处理技术和通信手段的飞速发展,信息隐藏和隐藏分析技术在信息安全中的作用越来越受到人们的关注。其中,音频隐写术是一种将秘密信息隐藏在普通的、非秘密的、可运行的音频文件中的技术。

[0003] 现有的音频隐写术主要是通过音频的时域特征设计算法,使用生成载体修改向量的方式达到隐写的目的。然而,该方法容易导致网络模型退化,不利于模型稳定训练,导致隐写的稳定性差。

发明内容

[0004] (一)解决的技术问题

[0005] 针对现有技术的不足,本发明提供了一种端到端音频隐写方法、系统、存储介质及电子设备,解决了现有的音频隐写方法稳定性差的技术问题。

[0006] (二)技术方案

[0007] 为实现以上目的,本发明通过以下技术方案予以实现:

[0008] 第一方面,本发明提供一种端到端音频隐写方法,采用生成对抗网络预先构建编码器和隐写分析器,根据编码器预先构建解码器,所述端到端音频隐写方法包括:

[0009] S1、获取秘密音频和载体音频,并通过预先训练的编码器对秘密音频和载体音频进行处理,输出载密音频;

[0010] S2、通过解码器对载密音频进行解密处理,输出秘密音频的估计音频;

[0011] 其中,通过循环自编码器进行生成对抗网络预训练,确定编码器中特征提取模块和特征还原模块的参数。

[0012] 优选的,所述特征提取模块用于提取并联合秘密音频时间依赖特征和载体音频时间依赖特征,得到时间依赖特征;

[0013] 所述特征提取模块包括依次连通的6个Convblock和1个拼接层,其中,第一个Convblock的输入通道数1,输出通道数64和卷积核大小为3×3,第二个Convblock的输入通道数64,输出通道数64和卷积核大小为1×3,第三个Convblock的输入通道数64,输出通道数128和卷积核大小为1×3,第四个Convblock的输入通道数128,输出通道数128和卷积核大小为1×3,第五个Convblock的输入通道数128,输出通道数128和卷积核大小为1×3;第六个Convblock的输入通道数256,输出通道数256和卷积核大小为1×3。

[0014] 优选的,所述编码器还包括特征嵌入模块,所述特征嵌入模块用于高维展开时间依赖特征,并进行秘密特征的嵌入,得到嵌入秘密音频特征的载密融合特征;

[0015] 所述特征嵌入模块包括依次连通的8个mixblock,8个mixblock的卷积核大小均为

3×3,其中,第一个mixblock的输入通道数512,输出通道数576,第二个mixblock的输入通道数576,输出通道数640,第三个mixblock的输入通道数640,输出通道数768,第四个mixblock的输入通道数768,输出通道数1024,第五个mixblock的输入通道数1024,输出通道数768,第六个mixblock的输入通道数768,输出通道数576,第七个mixblock的输入通道数576,输出通道数512,第四个mixblock的输入通道数512,输出通道数256。

[0016] 优选的,所述特征还原模块用于对载密融合特征进还原,输出载密音频;

[0017] 所述特征还原模块包括依次连通的6个Transblock,其中,前五个Transblock的卷积核为1×3,第六个Transblock的卷积核为3×3,第一个Transblock的输入通道数256,输出通道数256,第二个Transblock的输入通道数256,输出通道数128,第三个Transblock的输入通道数128,输出通道数128,第四个Transblock的输入通道数128,输出通道数64,第五个Transblock的输入通道数64,输出通道数64,第六个Transblock的输入通道数64,输出通道数1。

[0018] 优选的,所述隐写分析器包括依次连通的4个Convblock、3个Linearblock和一层softmax层。

[0019] 优选的,所述解码器中包括第二特征提取模块和第二特征还原模块,所述第二特征提取模块通过共享编码器中特征提取模块的网络参数,第二特征提取模块的结构、参数和编码器中特征提取模块中的结构、参数保持一致。

[0020] 优选的,所述编码器、隐写分析器和解码器训练过程中的损失函数包括:

[0021] $L_S = {}_x\log(S(C)) + (1-x)\log(1-S(C'))$

[0022] $L_D = \mathrm{Distortion}(M,M')$

[0023] $L_E = \lambda_1(\mathrm{Distortion}(C,C')) + \lambda_2 L_S + \lambda_3 L_D$

[0024] $$\mathrm{Distortion}(C,C') = 1 - \frac{\sum y_i' y_i}{\sum (y_i'^2 + y_i^2) - \sum y_i' y_i}$$

[0025] 其中,L_E表示编码器的损失;L_D表示解码器的损失;L_S表示隐写分析器的损失;λ_1、λ_2、λ_3分别表示编码器、隐写分析器、解码器的损失所占的权重系数;S(C)表示被隐写分析器S识别为载体音频的概率,S(C′)表示被识别为载密音频的概率;x表示隐写分析器的标签,将编码器产生的载密音频标签为1,将原始的载体音频标签为0;$y = \{y_1, y_2, \cdots, y_i, \cdots, y_n\}$表示时域载体音频,$y' = \{y_1', y_2', \cdots y_i' \cdots, y_n'\}$表示时域载密音频。

[0026] 第二方面,本发明提供一种端到端音频隐写系统,采用生成对抗网络预先构建编码器和隐写分析器,根据编码器预先构建解码器,该端到端音频隐写系统包括:

[0027] 加密模块,用于获取秘密音频和载体音频,并通过预先训练的编码器对秘密音频和载体音频进行处理,输出载密音频;

[0028] 解码模块,用于通过解码器对载密音频进行解密处理,输出秘密音频的估计音频;

[0029] 其中,通过循环自编码器进行生成对抗网络预训练,确定编码器中特征提取模块和特征还原模块的参数。

[0030] 第三方面,本发明提供一种计算机可读存储介质,其存储用于端到端音频隐写的计算机程序,其中,所述计算机程序使得计算机执行如上述所述的端到端音频隐写方法。

[0031] 第四方面,本发明提供一种电子设备,包括:

[0032] 一个或多个处理器,存储器,以及一个或多个程序,其中所述一个或多个程序被存

储在所述存储器中,并且被配置成由所述一个或多个处理器执行,所述程序包括用于执行如上述所述的端到端音频隐写方法。

[0033] (三)有益效果

[0034] 本发明提供了一种端到端音频隐写方法、系统、存储介质及电子设备。与现有技术相比,具备以下有益效果:

[0035] 本发明采用生成对抗网络预先构建编码器和隐写分析器,根据编码器预先构建解码器,该方法包括:获取秘密音频和载体音频,并通过预先训练的编码器对秘密音频和载体音频进行处理,输出载密音频;通过解码器对载密音频进行解密处理,输出秘密音频的估计音频;其中,通过循环自编码器进行生成对抗网络预训练,确定编码器中特征提取模块和特征还原模块的参数。本发明通过循环自编码器进行生成对抗网络预训练,确定编码器中特征提取模块和特征还原模块的参数,且基于生成对抗网络框架设计了端到端的隐写算法,不仅避免了因为STFT不匹配导致的秘密信息提取失败问题,同时取消了载体音频的修改向量,使编码器直接生成载密音频,从而达到降低模型的训练难度并提高模型性能的目的,有效解决了现有的音频隐写方法稳定性差的技术问题。

附图说明

[0036] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0037] 图1为本发明实施例中编码器、解码器以及隐写分析器的结构示意图;

[0038] 图2为编码器中的特征提取模块的结构示意图;

[0039] 图3为编码器中的特征嵌入模块的结构示意图;

[0040] 图4为编码器中的特征还原模块的结构示意图;

[0041] 图5a、5b为失真度约束随幅值y_k和修改幅度δ的变化示意图;

[0042] 图6a、6b为同一段音频的各音频向量的时域信号和时间依赖特征的均值和方差;

[0043] 图7a、7b为隐写前后的频谱图对比,其中图7a为嵌入信息前的频谱图,图7b为嵌入信息后的频谱图。

具体实施方式

[0044] 为使本发明实施例的目的、技术方案和优点更加清楚,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0045] 本申请实施例通过提供一种端到端音频隐写方法、系统、存储介质及电子设备,解决了现有的音频隐写方法稳定性差的技术问题,实现提高音频隐写算法的不可感知性、抗检测性和秘密信息的提取准确率。

[0046] 本申请实施例中的技术方案为解决上述技术问题,总体思路如下:

[0047] 现有的音频隐写方法主要包括基于时域特征设计的算法和基于短时傅里叶STFT

特征设计的算法,然而,这两个方法存在以下缺陷:

[0048] 1)以时域特征设计的算法,使用生成载体修改向量的方式达到隐写的目的,容易导致网络模型退化问题,不利于模型稳定训练,导致隐写的稳定性差,且会降低隐写算法的性能。2)基于短时傅里叶STFT特征所设计的音频隐写算法,容易受STFT失配问题的影响导致秘密信息的提取失败。

[0049] 为了克服上述缺陷,本发明实施例设计通过循环自编码器进行生成对抗网络预训练,确定编码器中特征提取模块和特征还原模块的参数,且基于生成对抗网络框架设计了端到端的隐写算法,降低模型的训练难度并提高模型性能的目的,有效解决了现有的音频隐写方法稳定性差的技术问题。

[0050] 为了更好的理解上述技术方案,下面将结合说明书附图以及具体的实施方式对上述技术方案进行详细的说明。

[0051] 本发明实施例提供一种端到端音频隐写方法,采用生成对抗网络预先构建编码器和隐写分析器,根据编码器预先构建解码器,该方法包括:

[0052] S1、获取秘密音频和载体音频,并通过预先训练的编码器对秘密音频和载体音频进行处理,输出载密音频;

[0053] S2、通过解码器对载密音频进行解密处理,输出秘密音频的估计音频;

[0054] 其中,通过循环自编码器进行生成对抗网络预训练,确定编码器中特征提取模块和特征还原模块的参数。

[0055] 本发明实施例通过循环自编码器进行生成对抗网络预训练,确定编码器中特征提取模块和特征还原模块的参数,且基于生成对抗网络框架设计了端到端的隐写算法,不仅避免了因为STFT不匹配导致的秘密信息提取失败问题,同时取消了载体音频的修改向量,使编码器直接生成载密音频,从而达到降低模型的训练难度并提高模型性能的目的,有效解决了现有的音频隐写方法稳定性差的技术问题。

[0056] 在本发明实施例中,采用生成对抗网络构建编码器E、解码器D以及隐写分析器S,其结构如图1所示。在训练阶段,编码器接收秘密音频M和载体音频C,输出载密音频C'。解码器旨在从载密音频C'中解码出秘密音频的估计音频M'。隐写分析器在框架中扮演人类的视角,负责鉴别音频被隐藏了秘密信息的概率P。编码器E为了达到欺骗隐写分析器S的目的,最终输出的C'要和C尽可能的相似,引入隐写分析器的目的是提高隐写的安全性。

[0057] 编码器中包括特征提取模块、特征嵌入模块和特征还原模块。特征提取模块首先从载体音频和秘密音频信号中提取得到时间依赖性特征,然后特征嵌入模块将特征从通道维扩展,再进行融合得到嵌入秘密音频的融合特征,最终经过特征还原模块将融合特征还原为载密音频信号。

[0058] 通过编码器的特征提取模块,可以有效捕捉音频信号的中短期和长期的依赖性特征。具体来说分为两个方面,通过向量内卷积和向量间卷积提取时间依赖性;通过低维表征降低时间分辨率提高特征的数值稳定性。特征提取模块结构如图2所示,包括依次连接的6个ConvblockConvblock和1个拼接层concat,其中,Convblock(m,n,k)各参数分别表示输入通道数m,输出通道数n和卷积核大小k,mixblock和transblock的参数含义相同。载体音频C和秘密音频M经过预处理划分为等长的音频向量。特征提取的第一层,设置卷积核大小为(3×3),相邻向量之间进行卷积,不进行降维操作;特征提取的后五层,卷积核设置为(1×3),

音频向量内进行卷积,同时降低分辨率为原来的一半。然后将提取的载体音频特征和秘密音频特征并联作为时间依赖特征(Time-feature)输入到特征嵌入模块。

[0059] 特征嵌入模块的作用是高维展开时间依赖特征(Time-feature),然后进行秘密特征的嵌入,网络结构如图3所示。Time-feature输入到特征融合网络,经过4层Mixblock将通道扩展到1024维,再经过4层Mixblock,融合通道维得到嵌入秘密音频特征的载密融合特征(Mix-feature)。

[0060] 载密融合特征可以被特征还原模块还原为音频数据,特征还原模块的网络结构如图4所示。为了保证输出音频保持和输入音频具有相同的分辨率,特征还原模块和特征提取模块的网络层需要一一对应(通道维的增加和减少相对应,数据的下采样和上采样相对应)。特征还原模块包括6个Transblock,前五层卷积核大小为(1×3),卷积向量内的特征,降低通道维,提升分辨率,最后一层的卷积核大小设置为(3×3),卷积相邻向量间的特征输出载密音频。

[0061] 解码器和隐写分析器的网络结构如表1所示。解码器的网络结构和去除特征嵌入模块的编码器相同,即解码器中包括特征提取模块和特征还原模块,解码器中的特征提取模块通过共享编码器中的特征提取模块的网络参数可以使解码器提取的特征和编码器保持一致,这样可以加速模型的训练,提高隐写性能。解码器的特征还原模块需要从载密特征中还原秘密音频,不需要与编码器共享参数。隐写分析器的任务是鉴别输入音频是否藏有秘密信息,并输出藏有秘密信息的概率。因此,首先经过Convblock提取深度特征,这里为了减少参数量,设置Convblock的步长为3,然后经过三个Linearblock将数据降为2维,经过softmax层输出预测概率。

[0062] 表1解码器、隐写分析器模型结构

[0063]

	解码器	隐写分析器
1	**input**	**input**
2	Convblock(1, 64)	Convblock(1, 4)
3	Convblock(64, 64)	Convblock(4, 8)
4	Convblock(64, 128)	Convblock(8, 16)
5	Convblock(128, 128)	Convblock(16, 32)
6	Convblock(128, 256)	Linearblock
7	Convblock(256, 256)	Linearblock
8	Transblock(256,256)	Linearblock
9	Transblock(256, 128)	softmax
10	Transblock(128,128)	**output**
11	Transblock(128, 64)	
12	Transblock(64,64)	
13	Transblock(64, 1)	
14	**output**	

[0064]

[0065] 在编码器、解码器和隐写分析器的训练过程中,对其损失函数进行优化。具体如下:

[0066] 生成对抗网络属于较难训练的一类网络,其在隐写领域的联合训练过程中,总共

包括三类损失:编码器E损失L_E、解码器D损失L_D和隐写分析器S损失L_S,如式(1)-(3)所示。隐写分析器判别输入音频的是否被嵌入信息,其损失属于分类损失,故采用常用的交叉熵分类损失。解码器提取准确率高的秘密音频,一般使用能衡量失真度的一类Distortion函数作为损失函数。编码器的损失由三部分组成,除Distortion函数外,还包括解码器和隐写分析器的损失,其中λ_1、λ_2、λ_3分别表示三者损失函数所占的系数。

[0067] $L_S=_x\log(S(C))+(1-x)\log(1-S(C'))$ (1)

[0068] $L_D=\mathrm{Distortion}(M,M')$ (2)

[0069] $L_E=\lambda_1(\mathrm{Distortion}(C,C'))+\lambda_2L_S+\lambda_3L_D$ (3)

[0070] 其中,S(C)表示被隐写分析器S识别为载体音频的概率,S(C′)表示被识别为载密音频的概率。x为隐写分析器的标签,将编码器产生的载密音频标签为1,将原始的载体音频标签为0。

[0071] 现有的隐写算法中常用的约束失真度的函数有MSE、L-P范数和SNRloss,如式(4)-(6)所示。其中,信噪比SNR计算公式通常被用来衡量相似度,SNRloss则是使用信噪比的负值来达到约束失真的目的。在回归任务中,这些损失多是基于标签值与预测值之间的差异来计算损失loss,通过减小loss来达到预测的目的。对于隐写而言就是通过减小标签值与预测值之间的失真程度,从而提高嵌入秘密音频后的载密音频和载体音频的相似度。

[0072] $$\mathrm{MSE}=\frac{1}{n}\sum(y_i-y_i')^2 \quad (4)$$

[0073] $$L-P\text{范数}=\sqrt[p]{\sum(y_i-y_i')^p} \quad (5)$$

[0074] $$\mathrm{SNRloss}=-10log_{10}(\frac{\sum y_i^2}{\sum(y_i-y_i')^2}) \quad (6)$$

[0075] 其中,$y=\{y_1,y_2,\cdots,y_i,\cdots,y_n\}$表示时域载体音频,$y'=\{y_1',y_2',\cdots y_i'\cdots,y_n'\}$表示时域载密音频。

[0076] 然而,音频作为一种电磁波,其在不同时间段的能量大小上有很大的差异,上述失真函数对能量不均衡的问题考虑不够全面。基于时域信号的音频隐写算法应该符合大振幅优先原则,即在能量较强的时间段应该嵌入更多的信息,在能量较弱的时间段相对的应该嵌入信息少一些。本发明实施例使用广义Jaccard系数优化失真度Distortion函数,达到针对采样点振幅的大小智能自适应的调整失真约束,如式(7)所示。

[0077] $$\mathrm{Jaccard}=1-\frac{\sum y_i'y_i}{\sum(y_i'^2+y_i^2)-\sum y_i'y_i} \quad (7)$$

[0078] 为了嵌入秘密信息,假设隐写算法需要将时域音频向量y的第k位y_k修改为y_k',修改幅度为δ,即$y_k=y_k'+\delta$,计算得到音频向量的失真度约束如表2所示。

[0079] 表2MSE、L-P范数、SNRloss和Jaccard的失真度约束

[0080]

失真度函数	失真度约束
MSE	$\frac{1}{n}\delta^2$
L-P范数	$\lvert\delta\rvert$或δ
SNRloss	$-10\log_{10}\frac{\sum_{i=1}^{k-1}y_i^2+\sum_{i=k+1}^{n}y_i^2+y_k^2}{\delta^2}$
Jaccard	$\frac{\delta^2}{\sum_{i=0}^{k-1}y_i^2+\sum_{i=k+1}^{n}y_i^2+y_k^2-y_k\delta+\delta^2}$

[0081] 其失真度约束随幅值y_k和修改幅度δ的变化如图5a、图5b所示。MSE和L-P范数事实上对音频的每个采样点的约束相同,幅值的大小并不影响其失真度。与之相对的jaccard和SNRloss对幅值敏感,可以根据y_k的大小自适应的调整失真度约束。另一方面,SNR loss在δ接近于零时,失真度的趋近于无穷,这限制了音频的修改幅度,且容易导致梯度爆炸,不利于模型的训练。相对的Jaccard失真度随δ的变化相对平缓,可接受的δ范围更广,更利于音频隐写。

[0082] 通过训练好的编码器和解码器对音频进行加密和解密,具体如下:

[0083] 在步骤S1中,获取秘密音频和载体音频,并通过预先训练的编码器对秘密音频和载体音频进行处理,输出载密音频。具体实施过程如下:

[0084] S101、特征提取模块提取并联合秘密音频时间依赖特征和载体音频时间依赖特征,得到时间依赖特征。具体实施过程如下:

[0085] 由于音频具有很高的时间分辨率(如16kHz),这使得单个音频信号几乎不具有实际意义,必须与附近的音频信号甚至相隔更远处的信号一起构成声音。因此,在隐写算法中使用的隐写特征需要能很好的捕捉这种时间依赖关系。如时序音频$y=\{y_1,y_2,\cdots y_i\cdots,y_n\}$,其对应的特征z在采样点i处的特征$z_i$可以由提取函数f从音频时序信号$y_{i-n}$到$y_{i+n}$中提取。

[0086] $z_i=f(y_{i-n},\ldots,y_i,\ldots y_{i+n})$

[0087] 为了充分的捕捉音频信号的中短时依赖和长期依赖,本发明实施例通过预先构建的编码器中的特征提取模块有效提取音频信号的时间依赖特征,时域信号按固定长度划分为音频向量,堆叠为输入矩阵送入特征提取模块。使用不同规格的卷积核(3×3)和(1×3),提取音频向量之间和向量内的时间依赖特征。此外,在提取特征的过程中,通过不断降低每一通道维的数据维度,来表征高分辨率的时间信号。降维的另一个好处是可以增加数值的稳定性,分别计算来自同一段音频的各音频向量的时域信号和时间依赖特征的均值和方差,其结果如图6a、6b所示。直接使用时域信号,其特征的均值和方差变化剧烈,数值偏小,不利于模型的训练。而降维的时间依赖特征的均值和方差稳定性更高,数值区间更加合理。

[0088] 此外,时间依赖特征还具有良好的可修改性,可修改性即嵌入信息后对特征的影响越小越好。频谱图是一种常用的分析信号特性尤其是频域特性的工具,在等长音频嵌入的情况下,对比嵌入前后的频谱图没有明显的差异,如图7a、7b所示,说明时间依赖性特征

具有很好的可修改性。

[0089] S102、通过预先训练的编码器中的特征嵌入模块对时间依赖特征进行处理,得到嵌入秘密音频特征的载密融合特征。

[0090] S103通过预先训练的编码器中的特征还原模块对载密融合特征进行处理,输出载密音频。

[0091] 在步骤S2中,通过解码器对载密音频进行解密处理,输出秘密音频的估计音频。具体实施过程如下:

[0092] S201、通过解码器中的特征提取模块提取载密音频中的时间依赖特征。解码器中的特征提取模块通过共享编码器中的特征提取模块中的网络参数得到。

[0093] S202、通过解码器中的特征还原模块对载密音频中的时间依赖特征进行处理,还原秘密音频,输出秘密音频的估计音频。

[0094] 下面通过对比实验验证本发明实施例的有效性:

[0095] 该实验将从载密音频的不可感知性、秘密音频的提取以及抗隐写分析三个方面与时域模型CNN-based、TCN模型和基于频域特征的模型BNSNGAN进行对比验证本发明实施例所提方法(proposed)。实验选用开源流行的Librispeech数据集构建2s和10s音频测试数据集。

[0096] 表3给出了四种算法在不同音频时长下的信噪比SNR、客观等级差异ODG和均方误差MSE.SNR和ODG用来衡量载密音频的不可感知性,MSE衡量秘密信息的提取误差。从表中可以看出本发明实施例所使用的算法其SNR值高于28,ODG均值平均在-1.5,MSE值小于0.00018,均优于TCN、CNN-based和BNSNGAN隐写算法。

[0097] 表3不可感知性测试结果

[0098]

	算法	SNR	ODG	MSE
2s	proposed	28.66	-1.43	0.00018
	TCN	22.12	-2.67	0.00140
	CNN-based	24.67	-2.43	0.00045
	BNSNGAN	26.61	-1.51	0.00026
10s	proposed	29.31	-1.59	0.00015
	TCN	21.75	-2.81	0.00170
	CNN-based	24.69	-2.61	0.00051
	BNSNGAN	26.99	-1.65	0.00024

[0099] 表4给出了使用两类隐写分析器analyzer1和analyzer2对四种算法生成的载密音频的检测结果。使用准确率ACC、虚警率FPR和漏检率FNR统计隐写分析的检测结果。从表中可以看出,经过两类分析器检测,本发明实施例所提出的隐写方法其ACC比TCN、CNN-based和BNSNGAN更低,FPR和FNR都更高。表明本发明实施例的方法对隐写检测器具有更好的欺骗性,抗检测性能更好。

[0100] 表4隐写检测测试结果

[0101]

算法	分析器	ACC	FPR	FNR
proposed	analyzer1	0.6315	0.3654	0.3714
	analyzer2	0.6950	0.2896	0.2801
TCN	analyzer1	0.9340	0.0450	0.0850
	analyzer2	0.8950	0.0977	0.1119
CNN-based	analyzer1	0.7680	0.2287	0.2351
	analyzer2	0.7795	0.2085	0.2315
BNSNGAN	analyzer1	0.7180	0.2640	0.2973
	analyzer2	0.8745	0.1243	0.1266

[0102] 表5对比以MSE、L1范数、SNRloss和Jaccard系数为失真度衡量的损失函数在算法上的性能表现,基于Jaccard系数优化的损失函数在SNR和ODG上均远大于其它损失函数,提取误差MSE值上远小于其它结果,表明了基于Jaccard系数优化的损失函数对隐写性能提升有很大的帮助。

[0103] 表5不同失真度函数的实验结果

[0104]

	失真函数	SNR	ODG	MSE
2s	MSE	23.78	-2.13	0.00042
	L1	18.25	-2.89	0.00066
	SNRloss	13.36	-3.13	0.00122
	Jaccrad	28.66	-1.43	0.00018
10s	MSE	22.76	-2.24	0.00034
	L1	17.36	-2.91	0.00064
	SNRloss	12.58	-3.24	0.00114
	Jaccrad	29.31	-1.595	0.00015

[0105] 本发明实施例还提供一种端到端音频隐写系统,采用生成对抗网络预先构建编码器和隐写分析器,根据编码器预先构建解码器,该系统包括:

[0106] 加密模块,用于获取秘密音频和载体音频,并通过预先训练的编码器对秘密音频和载体音频进行处理,输出载密音频;

[0107] 解码模块,用于通过解码器对载密音频进行解密处理,输出秘密音频的估计音频;

[0108] 其中,通过循环自编码器进行生成对抗网络预训练,确定编码器中特征提取模块和特征还原模块的参数。

[0109] 可理解的是,本发明实施例提供的端到端音频隐写系统与上述端到端音频隐写方法相对应,其有关内容的解释、举例、有益效果等部分可以参考端到端音频隐写方法中的相应内容,此处不再赘述。

[0110] 本发明实施例还提供一种计算机可读存储介质,其存储用于端到端音频隐写的计算机程序,其中,所述计算机程序使得计算机执行如上述所述的端到端音频隐写方法。

[0111] 本发明实施例还提供一种电子设备,包括:

[0112] 一个或多个处理器;

[0113] 存储器;以及

[0114] 一个或多个程序,其中所述一个或多个程序被存储在所述存储器中,并且被配置

成由所述一个或多个处理器执行,所述程序包括用于执行如上述所述的端到端音频隐写方法。

[0115] 综上所述,与现有技术相比,具备以下有益效果:

[0116] 1、本发明实施例通过循环自编码器进行生成对抗网络预训练,确定编码器中特征提取模块和特征还原模块的参数,且基于生成对抗网络框架设计了端到端的隐写算法,不仅避免了因为STFT不匹配导致的秘密信息提取失败问题,同时取消了载体音频的修改向量,使编码器直接生成载密音频,从而达到降低模型的训练难度并提高模型性能的目的,有效解决了现有的音频隐写方法稳定性差的技术问题。

[0117] 2、本发明实施例提取音频的时间依赖特征捕捉时域信号的长短时依赖,从而获得适合用于音频隐写的特征并提高了特征的数值稳定性,避免了数据偏小和波动幅度剧烈导致的性能下降,解决了现有的音频隐写方法稳定性差的技术问题,实现提高音频隐写算法的不可感知性、抗检测性和秘密信息的提取准确率。

[0118] 3、使用广义Jaccard系数作为失真度量,优化生成对抗网络的损失函数,可以自适应调整约束,解决基于GAN模型的隐写算法训练慢、收敛差的缺陷。

[0119] 4、在编码器中设计特征嵌入模块,融合秘密信息和载体音频的深度特征,以载体音频特征为基础指导编码器输出自然度高、不宜检测的载密音频。避免了基于GAN模型的隐写算法在训练时需要大量样本、容易模式坍塌的问题。

[0120] 需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0121] 以上实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

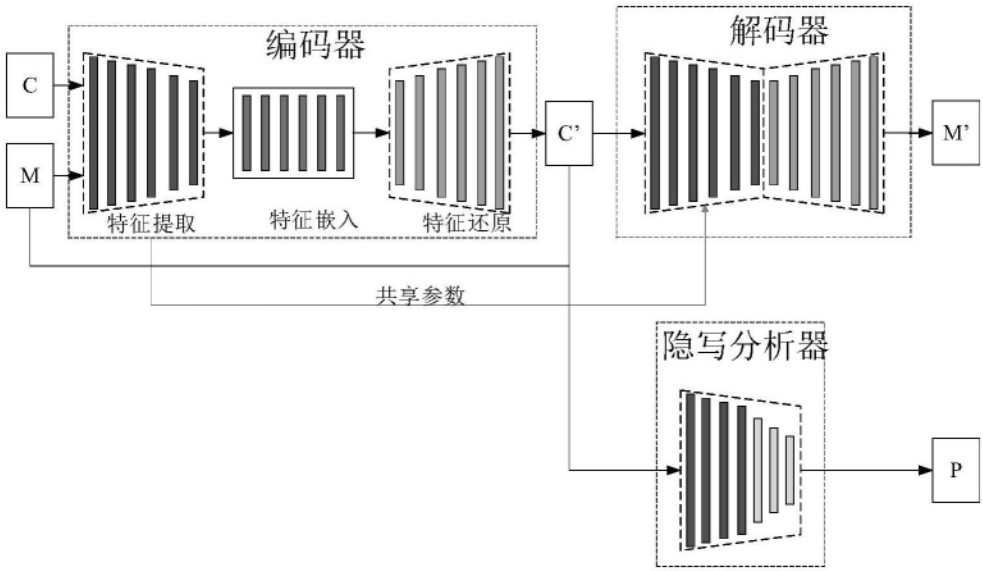

编码器
解码器
C
M
特征提取
特征嵌入
特征还原
C'
M'
共享参数
隐写分析器
P

图1

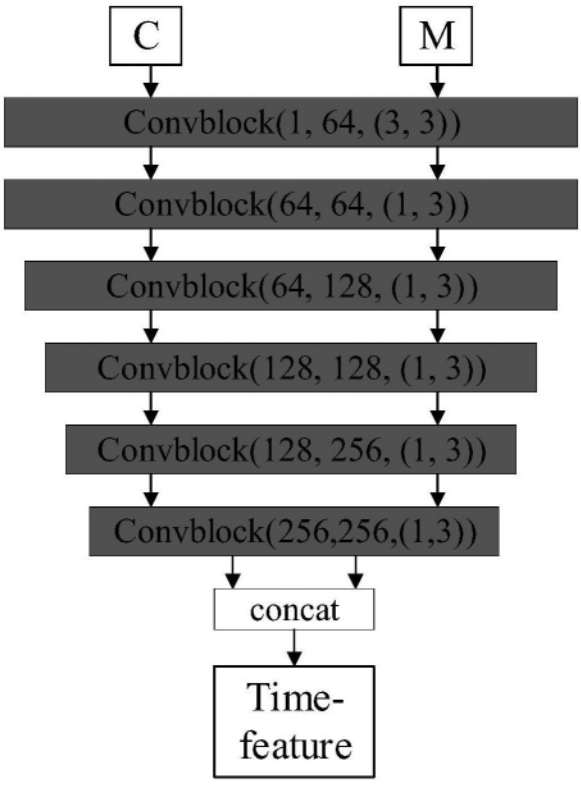

C
M
Convblock(1, 64, (3, 3))
Convblock(64, 64, (1, 3))
Convblock(64, 128, (1, 3))
Convblock(128, 128, (1, 3))
Convblock(128, 256, (1, 3))
Convblock(256,256,(1,3))
concat
Time-feature

图2

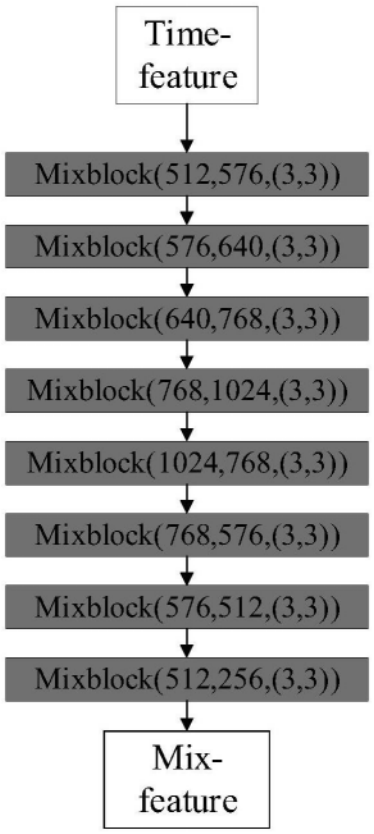

Time-feature
Mixblock(512,576,(3,3))
Mixblock(576,640,(3,3))
Mixblock(640,768,(3,3))
Mixblock(768,1024,(3,3))
Mixblock(1024,768,(3,3))
Mixblock(768,576,(3,3))
Mixblock(576,512,(3,3))
Mixblock(512,256,(3,3))
Mix-feature

图3

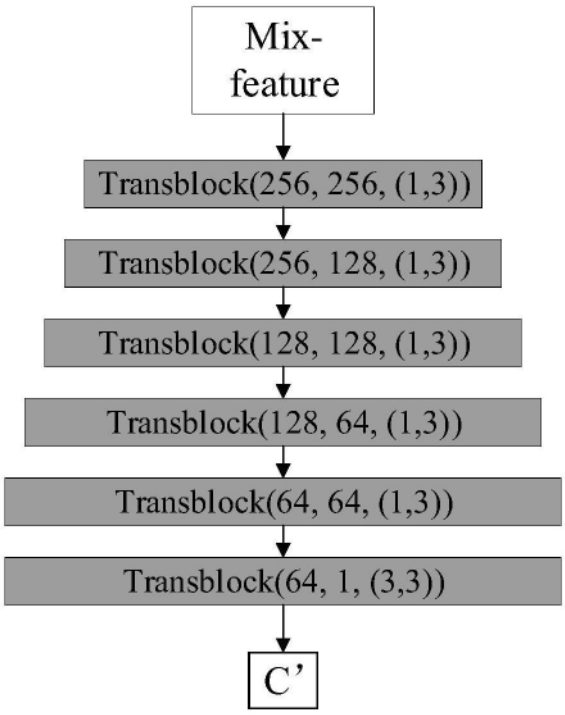

Mix-feature
Transblock(256, 256, (1,3))
Transblock(256, 128, (1,3))
Transblock(128, 128, (1,3))
Transblock(128, 64, (1,3))
Transblock(64, 64, (1,3))
Transblock(64, 1, (3,3))
C'

图4

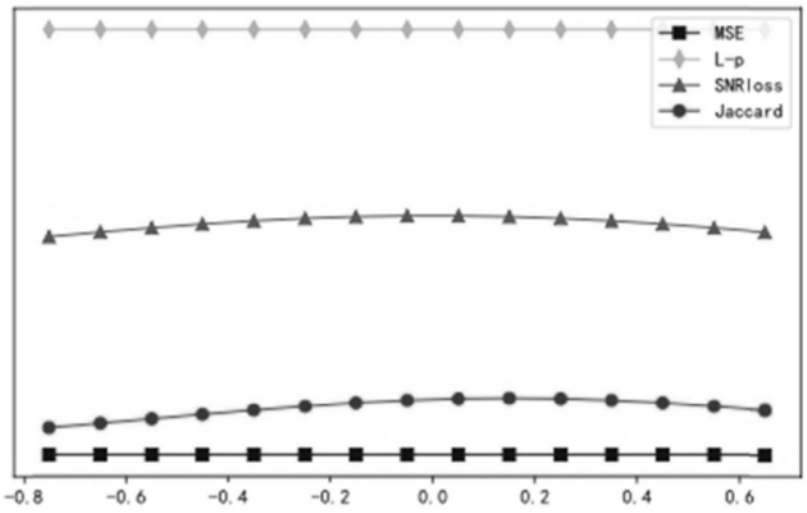

MSE
L-p
SNRloss
Jaccard
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6

图5a

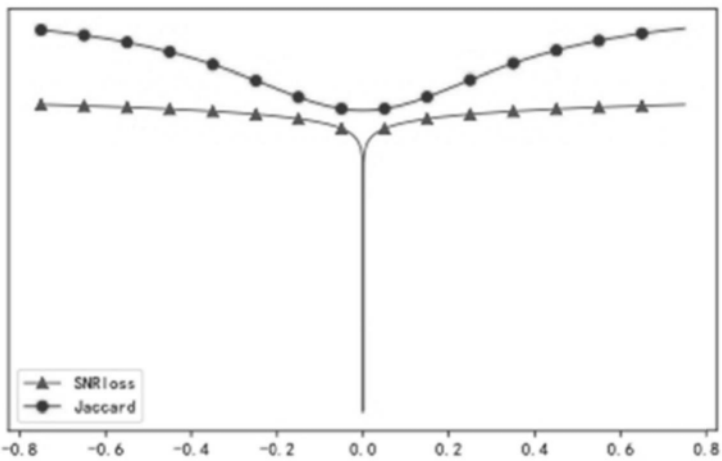

SNRloss
Jaccard
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8

图5b

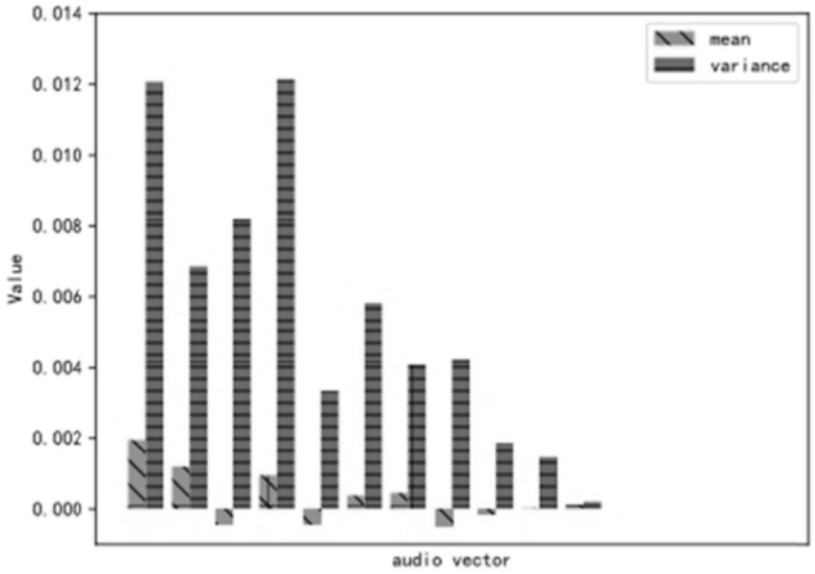

0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
Value
mean
variance
audio vector

图6a

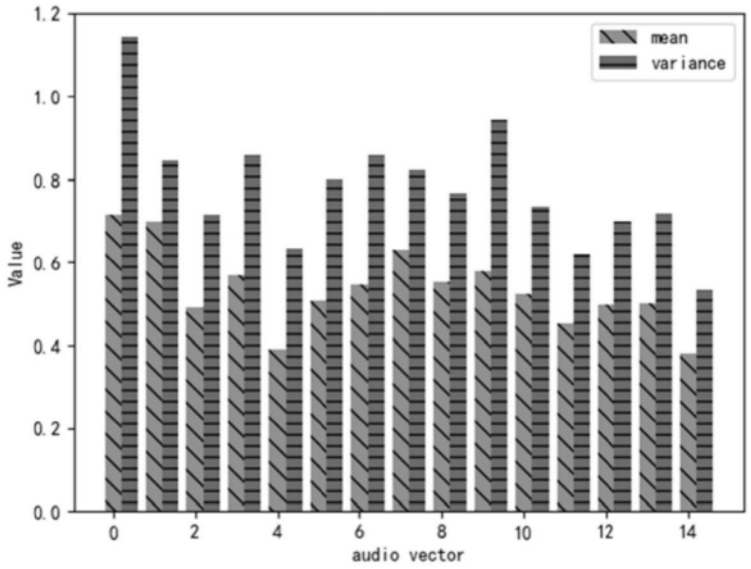

1.2
1.0
0.8
0.6
0.4
0.2
0.0
Value
mean
variance
0
2
4
6
8
10
12
14
audio vector

图6b

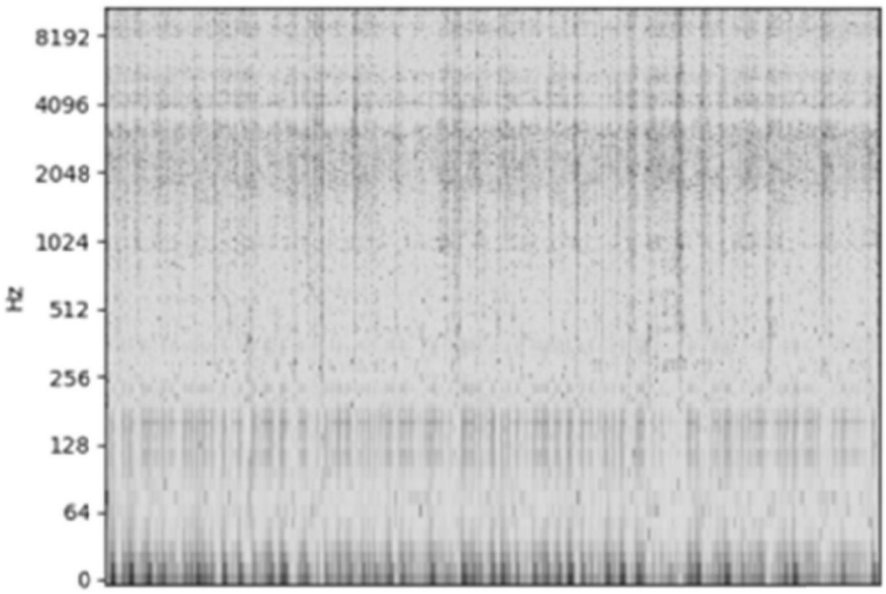

8192
4096
2048
1024
Hz
512
256
128
64
0

图7a

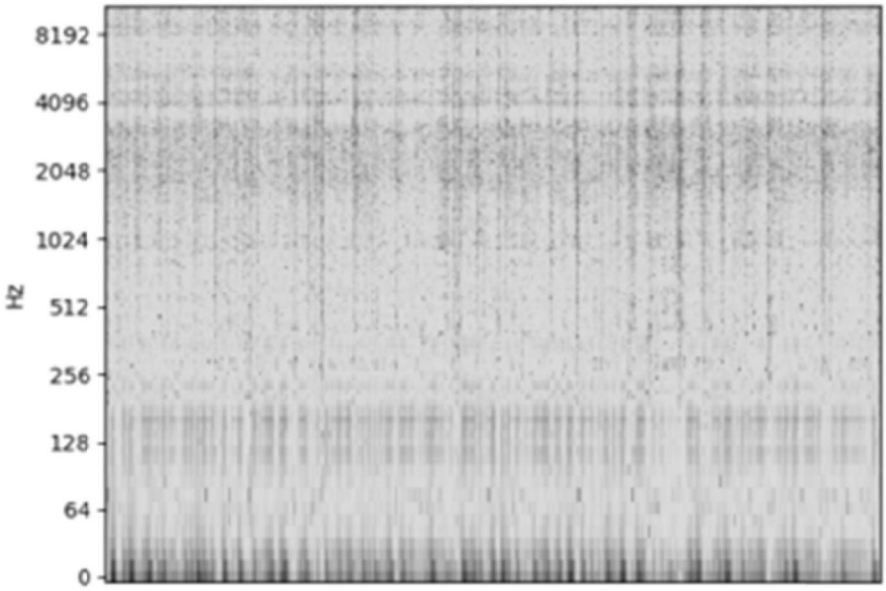

8192
4096
2048
1024
Hz
512
256
128
64
0

图7b